Recommendation Algorithms on the Web

ABSTRACT

In this paper, I will go over recommendation systems, which are algorithms used for content recommendation. Recommendation systems have become increasingly important on the Internet as more and more users find themselves on a handful of centralized websites that use such systems to help decide what content they will be shown. Using the recommendation system that provides the highest user satisfaction is critical for modern social media (and many other types of) sites to retain users. Therefore, exploring the current available algorithms and principles that make a successful recommendation system is a worthy endeavor.

1 INTRODUCTION

Recommendation systems have become increasingly important in recent years, especially in terms of social media, product recommendation on sites like Amazon or various streaming services, as well as for creating targeted advertisements. These systems have become an integral part of these companies' business models, with millions in potential revenue depending on how well these algorithms perform, while simultaneously having a massive effect on the experience of internet users (quite literally deciding what you see online). Yet most users have little to no idea how this type of system actually works, with "the algorithm" taking on a sort of magical reputation, whether its effect be good or bad. Throughout the rest of this paper we will go over the 4 main types of recommendation systems [1]: content-based, collaborative-filtering, knowledge based, and hybrid based. The hope is that after reading this users will have a better idea of how these algorithms tend to work and what kind of user information is useful for the companies to collect from you in order to build out these systems.

2 Types of Recommendation Systems

As mentioned previously, recommendation systems come in 4 main types. This section will give a brief overview of the types including how they generally work and what cases they're better suited for.

2.1 Knowledge-Based Recommendation Systems

We'll start off with one of the more basic and least "magic" of the recommendation systems, the knowledge-based recommendation system. This type of recommendation system is actually specifically useful for the problem where there might not be a lot of information about a given user [2]. Think of someone buying a very large, expensive item such as a new home. Most people buy homes extremely infrequently, so building up a user profile on a user's preferences in terms of homes is virtually impossible. This problem is known as the "cold-start problem" and knowledge-based recommendation systems are a good way to deal with it. These systems typically will require some type of user input at the start and then make some recommendations based on that. For example, for buying a home, a user might provide a cost range and general location for where they want to buy a home. The system can then start by recommending homes that match this criteria. Up until now, this "system" is basically just matching tags on its inventory with a user query. But the nice thing about this system is it can be built upon to do much more than that. For example, if users aren't finding something they like in the original search, the system could begin to recommend what it thinks are "similar" results in the hope of finding something adequate. In our home example, this might be houses just beyond the specified location or outside the specified price range. Additionally, this can be done using what the user has looked at rather than their query itself. If it's found that all the homes the user is looking into are within a certain size range, or have multiple bathrooms, the search can begin to narrow down provided results to only those matching what the user seems more interested in. Beyond that, a good recommendation system will give users a way to "tweak" the settings a bit as they see the results that are provided, as well as possibly making recommendations on what settings the user might be interested in specifying further based on user activity up until that point. It should be noted that unlike some other recommender systems, knowledge-based systems depend on expert domain knowledge (hence the name) to know which properties may be useful to search on and what further properties a user is likely to be interested in given what they've done so far [3]. They also are different in that they don't require a "profile" for a user storing the user's actions or preferences over a long period of time. These systems can quickly build up good recommendations over a single session.

2.2 Content-Based Recommendation Systems

The content-based recommendation system is a bit more in line with what people think of when they think of an "algorithm" that recommends things to them. This approach builds up a user profile for each user based on what items the user has said they like, or have just generally been looking at. Unlike the knowledge-based system, this algorithm does have initial difficulty making good recommendations in the beginning, but as a user's profile continues to be built up this approach can start to get powerful insights into the user's interests that might not be apparent using a knowledge-based system. By collecting user's ratings of various content items they come across, the system can begin to understand preferences the user has that they might not even be aware of, and recommend appropriately. This can be done by creating internal representations of the content items (as well as possibly the users) that can encode hidden or "latent" features of the content items [4]. It may then be able to find that users are interested in items that match these hidden features, or perhaps not-so-hidden features the user was unaware they were interested in, and generate new recommendations based on this. This system is nice in its simplicity. Basically as long as the service provides a way for users to rate the content they see (both directly or using indirect measures such as watch time) and has proper internal representations of the content items such that the features of items are taken into account in the representation such that from a set of items it can be seen in which way they are similar, these models can be built in a straightforward manner. Additionally, given users are effectively providing feedback on recommendations through further rating of the content, these models can be continually fine-tuned, including building more advanced machine learning models such as neural networks or transformers that are trained on this collected data and have been found to be better at finding hidden/latent similarities among content user rates highly, leading to even better performance in terms of recommending content that users will end up rating highly. Because these types of recommendation systems are straightforward at a high level they can also be extended to a number of use cases beyond what someone typically thinks a recommendation system can do, such as finding scientific papers [6] or discovering high quality art pieces [7].

2.3 Collaborative-Filtering Recommendation Systems

Up until now, all the recommendation systems covered have focused on one individual user. Knowledge-based recommendation systems use a single session building of a query to find the most relevant content, while content-based recommendation systems use the ratings of a user on certain content to recommend other content that is similar to the content the user has rated highly. These systems work well in many cases, but they miss out on the connections between users that are available in certain contexts such as followers or friends on various social media sites as well as just the general principle that more similar users are likely to want similar types of content. This is where a collaborative-filtering recommendation system can be of great use. Unlike other systems discussed that focus on one individual, collaborative-filtering recommendation systems build networks of users and can make decisions on what to recommend to a specific user based on their location inside this network. Similar to content-based approaches, this type of system needs to collect user opinions (or watch time, clicks, etc.) on certain types of content and have representations for the users and their content internally [5]. But it extends this idea by looking at users it considers “similar” to the given user. How this similarity is determined is a crucial point in the system. For certain situations like on a social media website this may be done in a straightforward manner by just assuming users that choose to link to each other as friends, followers, etc. are similar. In other situations where this information may not be available this can be done by grouping users by what type of content they look at/like. The hope being that if a user likes most of the same content as a group of other users, then they are likely to also like the other content that group likes. This makes sense, as there is a lot of content out there and it’s totally possible a user just hasn’t found some content they would very much enjoy. This is in fact one of the major advantages of this type of recommender system over the previously discussed systems. It is much better at helping users find totally new content they like because it uses information beyond just what the user themselves has been looking at and rating highly. Once the network of users is built, it is relatively straightforward to see how these algorithms might simply watch what other users in a user's network are interested in and recommend those things. In certain situations, this system can even be used to recommend new users to add to the user's network. For example, if a lot of people somebody is friends with on Facebook are friends with another individual, it would make sense to recommend that user as a new friend. This then continues to build on itself, incorporating this new friend's interests into the recommendations made to the original user. The system is not entirely without flaws however. For example, a user who is interested in an unpopular piece of content may still be unlikely to be recommended this content because, due to its unpopularity, other users in the user's network may not like this content , even if the user typically has their interests lined up with the other members of their network. That problem wouldn't exist in a purely content-based recommendation system that is only interested in what this user themselves likes. Additionally, this type of recommendation system has the downside that it incentivizes gathering an increased amount of information about users, as the system is now interested in understanding the user themselves in order to match them up with similar users rather than purely caring about what content the user has rated highly. Nonetheless, collaborative-filtering recommendation systems are very popular in situations where information to link users together are available (e.g. social media sites) and these types of recommendation systems still tend to do a better job at helping established users find new types of content than the previously discussed systems.

2.4 Hybrid Recommendation Systems

A hybrid recommendation system is simply a combination of the types of systems already discussed here. Typically, this involves combining a collaborative-filtering model with either a knowledge-based system or a content-based systems [1] in order to cover some of the problems that those systems are good at handling such as the cold-start problem that knowledge-based systems handle well or the issue of recommending unpopular items that content-based systems handle better than collaborative-filtering systems. The main task with a hybrid model is how the models should be combined. There are a wide variety of ways this can be done and no one methodology is clear cut as the "correct" way [9]. Those building these systems may have to experiment with different methods for making this happen. This can range from simply running both models at once and using some kind of weighted output of the different models, to incorporating one model into another by using one model's output as input to the other. Regardless of the method used, these hybrid models are extremely powerful in their flexibility and ability to cover a wide range of use cases effectively. Many recommendation systems for large platforms can be expected to use some type of hybrid model in the future if they are not already.

3 SUMMARY

Recommendation systems are playing an increasing role in people's lives, from content recommendation on large platforms, to even discovering scientific papers or artwork [6][7]. Due to this, it makes sense for an educated internet user to want to have some basic understanding of how these types of systems work, including both how the recommendations are done and what type of information these platforms are looking to gather from users in order to make these recommendations possible. This paper gives a brief overview of the 4 main types of recommendation systems, allowing users to gleam a basic idea of how these systems work, as well as giving them a jumping point if they decide they want further information on the topic.

4 REFERENCES

[1] Yuanzhe Peng. 2022. **A SURVEY ON MODERN RECOMMENDATION SYSTEM BASED ON BIG DATA**

[2] Ayesha Ameen. 2019. **Knowledge based Recommendation System in Semantic Web - A Survey**

[3] Robin Burke. 2000. **Knowledge-based recommender systems**

[4] Xavier Thomas. 2020 **Content-Based Personalized Recommender System Using Entity Embeddings**

[5] Weizhe Lin, Linjun Shou, Ming Gong, Pei Jian, Zhilin Wang, Bill Byrne, Daxin Jiang. 2022. **Transformer-Empowered Content-Aware Collaborative Filtering**

[6] Titipat Achakulvisut , Daniel E. Acuna , Tulakan Ruangrong, Konrad Kording. 2016 **Science Concierge: A fast content-based recommendation system for scientific publications**

[7] Antoine Fosset, Mohamed El-Mennaoui, Amine Rebei1, Paul Calligaro1, Elise Farge Di Maria1, Hélène Nguyen-Ban1, Francesca Rea1, Marie-Charlotte Vallade1, Elisabetta Vitullo1, Christophe Zhang1, Guillaume Charpiat, Mathieu Rosenbaum. 2022. **Docent: A content-based recommendation system to discover contemporary art**

[8] Dominik Kowald, Emanuel Lacic. 2022. **Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems**

[9] Erion C, Maurizio Morisio. 2019. **Hybrid Recommender Systems: A Systematic Literature Review**